Applied Soft Computing 38 (2016) 518–529

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Identifying association rules of specific later-marketed products

Cheng-Hsiung Weng*

Department of Management Information Systems, Central Taiwan University of Science and Technology, Taichung 406, Taiwan, ROC

ARTICLE INFO

Article history:
Received 12 August 2014
Received in revised form
14 September 2015
Accepted 21 September 2015
Available online 20 October 2015

Keywords:
Data mining
Association rule
Later-marketed product
Marketed time
Sale association

ABSTRACT

Not all products are marketed at the same time. If item (x) is marketed much earlier than item (z) is, then item (x) is associated with higher support compared with itemset (xz). In this situation, itemset (xz) cannot satisfy the minimum support; the association rule, $x \Rightarrow z$, possesses low confidence. To create better marketing strategies, managers must understand the sale associations between (x) and (z) and use (x) to promote (z) to increase the sales of (z). However, using traditional approaches for identifying the sale associations between earlier-marketed items and later-marketed item is difficult. In this study, we propose a new algorithm for determining the association rules by precisely calculating the support values of association rules. The association rules, which consist of an atomic consequent and its antecedents, consider the first time the consequent and its antecedents occurring in transactions. Furthermore, a new measure, *TransRate*, was designed to prevent generating useless itemsets. Experimental results from survey data indicated that the proposed approach can facilitate identifying rules of interest and valuable associations among later-marketed products.

1. Introduction

Association rule mining leads to the discovery of associations and correlations between items in large transaction data sets. The determined associations and correlations facilitate business decision-making processes. Association rule mining is a valuable data mining approach that can uncover consumer purchasing behaviors from transaction databases [1]. Agrawal et al. [2] first introduced the problem of association rule mining and defined it as finding all the rules from transaction data that satisfy the minimum support and minimum confidence constraints. The two thresholds, minimum support and minimum confidence, are used to evaluate frequent itemsets and association rules, respectively.

Not all products can be marketed at the same time, and products marketed at a later period may be associated with relatively low support and confidence [3]. Hence, determining the association rule for later-marketed products whose support or confidence is lower than the minimum support and confidence thresholds is impossible. Therefore, we argue that each item's marketed time should be considered when using association rule mining algorithms. Then the association rules can precisely present the associations within the itemsets. When focusing on later-marketed items, we should precisely calculate the support of the specific later-marketed consequent's antecedents, which were marketed early. This issue is illustrated in the following example.

Assume we have three products named x, y, and z (Fig. 1). Each product's marketed time is distinct. From Fig. 1, the following information can be derived:

(1) count(xyz) = 20 is the count of transactions in which products xyz occur since June,
(2) count(xy) = 10 + 10 = 20 is the count of transactions in which products xy occur since June,
(3) count(xy) = 10 + 10 + 10 + 10 = 40 is the count of transactions in which product xy occur since April,
(4) count(x) = 40 is the count of transactions in which product x occurs since April, and
(5) count(x) = 50 is the count of transactions in which product x occurs since March.

Product y is not yet sold in the market in March. Consumers bought product x without product y before April. Furthermore, products x and y were bought together after April. According to the preceding explanations, we know that:

(1) $conf(x \Rightarrow y) = 40/50 = 80\%$ since March,
(2) $conf(x \Rightarrow y) = 40/40 = 100\%$ since April,
(3) $conf(xy \Rightarrow z) = 20/40 = 50\%$ since April, and
(4) $conf(xy \Rightarrow z) = 20/20 = 100\%$ since June.

* Tel.: +886 4 22391647.
E-mail address: chweng@mgt.ncu.edu.tw

http://dx.doi.org/10.1016/j.asoc.2015.09.047

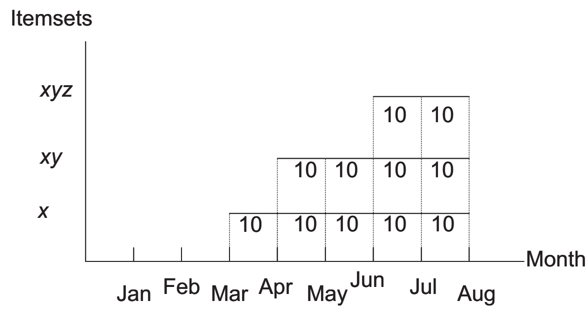

Fig. 1. Accumulative count for the later-marketed consequent product.

The rule [$conf(xy \Rightarrow z) = 20/20 = 100\%$ since June] seems more precise because product z is not marketed before June. We seek to precisely understand the association between the antecedent (xy) and the later-marketed consequent (z). If we want to uncover the association between the antecedent (xy) and the later-marketed consequent (z), we should consider the first time that both the antecedent and the later-marketed consequent product occur in the transactions. This is the beginning time from which to calculate the supports of products xyz and xy when generating association rule $xy \Rightarrow z$. This study aimed to identify the association rules of interest by precisely calculating the support values of association rules, which consist of an atomic consequent (later-marketed item) and its antecedents (earlier-marketed items), by considering the first time the atomic consequent and its antecedents occur in the transactions. Moreover, by focusing on the rules with the atomic item in the consequent, we can precisely realize the association of the atomic consequent and its antecedents; that is, which antecedents can be used to promote the atomic consequent when designing marketing strategy.

The rest of this paper is organized as follows: Section 2 reviews related work. The theoretical framework is presented in Section 3. The proposed algorithm and an example are illustrated in Section 4. Section 5 presents the demonstration of the effectiveness of the proposed algorithm, which was conducted using survey data. Conclusions and future work are presented in Section 6.

2. Related work

The main purpose of this study was to determine the specific later-marketed consequent association rules from transaction data. In this section, we explore the problem of association rule mining and introduce techniques related to association rules. Association rule mining and consequent association rule studies are reviewed.

Association rule mining is a data mining approach that can be used to determine consumer purchasing behaviors from transaction databases. Frequent pattern mining and association rule mining entail extracting regularities, correlations, and dependencies from databases. Frequent pattern mining reveals the intrinsic properties of data sets and is the foundation for association rule mining. Association rules are interesting, unexpected association relationships among attributes that satisfy minimum support and confidence in a database [1]. Agrawal et al. [2] first introduced the problem of association rule mining and defined it as finding all the rules from transaction data that satisfy the minimum support and the minimum confidence constraints. An association rule mining algorithm comprises two steps: (1) generate all frequent itemsets that satisfy the minimum support, and (2) generate all association rules that satisfy the minimum confidence from the already discovered frequent itemsets.

The Apriori algorithm has been widely used to generate all the frequent itemsets contained in a transaction database. Frequent itemsets mining algorithms are classified into three categories according to the data types they can handle: nominal/Boolean data [2], ordinal data [4], and quantitative data [5–7].

In contrast to single-level association rule mining, multilevel association rule mining emphasizes on identifying associations at multiple conceptual levels [8]. Because quantitative data are common in practical databases, a natural extension is to derive association rules by using quantitative data. To determine such association rules, the value of a quantitative attribute can be partitioned into a set of intervals to apply traditional algorithms for nominal data. These partitioning methods can be categorized into two major approaches, crisp [9] and fuzzy partitions [10,11].

Mining frequent itemsets in association rule mining is crucial [2]. Most of the frequent itemset mining algorithms are improved or derivative algorithms that are based on Apriori [12] or FP-growth [13]. Moreover, efficient methods for mining frequent itemsets have been proposed such as H-mine [14] and Index-BitTableFI [15]. However, most of these algorithms focus on improving the efficiency in frequent itemset mining processes rather than mining specific itemsets such as specific later-marketed items. To describe the differences between our and prior studies in frequent itemset mining, we provide comparison data in Table 1.

Identifying association or correlation relationships of interest facilitates business decision-making processes [16,17]. Na and Sohn [22] proposed an approach for predicting changes in the Korea Composite Stock Price Index on the basis of time series data of various interrelated world stock market indices. Huang et al. [18] introduced an association-rule-mining-based approach for determining interesting resource allocation rules from event logs. Chen et al. [19] applied the properties of propositional logic and proposed an algorithm for mining highly coherent rules; the derived association rules are more meaningful and reliable. Le et al. [20] proposed a new algorithm to remove sensitive knowledge from a released database according to the intersection lattice of frequent itemsets.

In commercial environments, commodities are usually introduced to the market over time; not all products are promoted to consumers simultaneously. In general, the time factor provides a higher support to previous marketed items, which means that later-marketed items have inferior support. In addition, because the support or confidence does not satisfy the minimum support and the minimum confidence constraints, it is unavailable for the creation of the association rules [3]. In the study by Chiang et al., the formula of support is equal to the confidence to extract the transaction of the interested itemsets. However, with this method, we cannot understand the difference in confidence values of the association rules comprising the same frequent itemsets. Moreover, they did not consider the first time of each item occurring in the transaction set. Table 2 shows a comparison of disjunctive consequent association rules of Chiang et al. [3] and those used in the current study.

Sequential pattern mining involves analyzing data represented as sequences (a sequence contains sorted sets of items). Compared with the association rule problem, a study of such data provides “inter transaction” analysis [21]. Sequential pattern mining analyzes data from a database of customer transactions; each transaction possesses the following characteristics: sequence ID or customer ID, transaction time, and transaction item. However, we required only the transaction time and transaction item. This is because sequential pattern mining approaches entail examining the sequence of items bought by a customer, whereas our study focused on identifying item associations without considering customer ID. Table 3 shows a comparison between sequential pattern mining and the approach used in this study.

In this study, we precisely calculated the support values of antecedent products (focusing on the later-marketed consequent product) to identify the association rules of interest and further